Governance risk mitigation, Commitment 18: Whistleblowing protections

On behalf of **Whistleblower-Netzwerk e.V.** and **OAISIS**, we are pleased to submit our comments on Commitment 18 of the **Second Draft General-Purpose AI Code of Practice**. Whistleblower-Netzwerk has been actively involved in providing expert opinions and participating in parliamentarian hearings regarding the **Directive (EU) 2019/1937** and its implementation in Germany and other laws and standards related to whistleblower protection. Our contributions have focused on the necessary legal and institutional changes to ensure that whistleblowers are adequately protected and supported. **OAISIS** is focused on supporting concerned insiders in the AI sector, and particularly well-versed in their needs.

Whistleblowers play a crucial role in uncovering misconduct and systemic risks, enabling early intervention to prevent harm and damages. To fulfill this vital role, whistleblowers require accessible and trustworthy reporting channels, a sense of impact, legal clarity, and robust protection against retaliation, as we know from our many years of advising whistleblowers. They also need independent advice and support to navigate the complexities of reporting. Reiterating key obligations from Directive (EU) 2019/1937 in the General Purpose AI Code of Practice is an important step, especially given the uneven implementation of the Directive across EU Member States. However, to further lower barriers for whistleblowers and to ensure that reported risks are effectively addressed, we propose that "*The commitment should be substantially edited and/or further clarified*" with the following amendments. (Our proposals are underlined, the type of change indicated is in green brackets, and the explanation of the changes in italic.)

[…] The following provisions are intended to highlight the most significant obligations and do not constitute an exhaustive list of all responsibilities under Directive (EU) 2019/1937:

- establishing secure and confidential internal reporting channels that allow for both written and oral reporting. Signatories are encouraged to provide various reporting channels, including one allowing anonymous follow-up communication with the whistleblower. (Extension of scope)
  Explanation: Anonymous reporting channels enable whistleblowers to build trust with the reporting body before disclosing their identity, thereby encouraging them to come forward. Systematic surveys of organizations have shown that implementing anonymous whistleblowing channels does not promote a culture of denunciation. The benefits of anonymous reporting channels significantly outweigh their (manageable) costs.
- designating an impartial person or department to handle reports. Reporting channels may be operated internally by a designated person or department or provided externally by a third party. Large organizations are encouraged to offer both an external third-party option (e.g., an ombudsperson) and an internal contact person. (Extension of scope)
  Explanation: Some whistleblowers may have greater trust in employees familiar with the internal context, while others may prefer an external party perceived as more independent. Offering a diverse range of reporting channels, including both internal and external options, enhances trust and encourages whistleblowers to come forward with critical information.
- acknowledging receipt of a report within seven days and giving substantiated feedback to the reporting persons about the action envisaged or taken as follow-up, for instance, closure of the procedure based, or launch of an investigation, and, as far

as possible, its findings and any measures taken to address the issue raised. (**Major** clarification)
Explanation: Whistleblowers often act out of a strong moral imperative to bring about change. However, this often leaves them in a vulnerable position. Open and transparent communication regarding the handling of their reports, including follow-up measures taken, is crucial for building trust and encouraging them to continue reporting misconduct. This objective is undermined by countries like Germany, which impose limited follow-up duties on authorities. These duties typically end as soon as the information is transferred to another authority competent to investigate the breach.

- providing clear and comprehensive information about the circumstances under which whistleblowers are protected, grievances that fall outside the material scope of protection, the reporting procedures, and the internal and external reporting channels available to them. This should also include details on additional resources for advice and support, such as civil society organizations or joint bodies established by the signatories. (Clarification)
  Explanation: Clear guidance on the circumstances under which they are protected from retaliation encourages whistleblowers to come forward with concerns without fear of negative consequences.
- explicitly informing individuals working for them of their right to choose between internal and external reporting channels, including the recommendation to use both channels in the case of time sensitive risks where immediate action is necessary to prevent significant harm. (**Major** clarification and extension of scope)
  Explanation: Relying solely on internal reporting processes in the case of time sensitive risks can lead to delays that may exacerbate the situation and hinder the ability to effectively address systemic risks
- protecting whistleblowers from all forms of retaliation, including dismissal, demotion, or harassment. This protection applies to all individuals who have encountered violations in a professional context, including current and former employees, contractors, and other stakeholders. Signatories are encouraged to extend protection from retaliation to individuals working for them, regardless of their location and including those outside the EU. (**Major** clarification and extension of scope)
  Explanation: This provision clarifies that the personal scope of the Directive (EU) 2019/1937 is not limited to employees. It also seeks to incentivize organizations to extend this protection to individuals working for them who may not fall within the scope of the Directive (EU) 2019/1937.
- protecting whistleblowers who had reasonable grounds to believe, at the time of reporting, that the information they provided was true and constituted a breach of the AI Act or the commitments outlined in this Code of Practice. (Clarification and minor extension of scope)
  Explanation: This provision clarifies the material scope of application, ensuring that also the commitments of the Code of Practice are covered, and emphasizes that whistleblowers only need to act in good faith.
- protecting public disclosures if the whistleblower has reported internally and/or externally, but no appropriate action has been taken within the timeframe referred to in the Directive (EU) 2019/1937 or if there is an imminent or manifest danger to the public interest, or if there is a risk of retaliation. Public disclosures may also be protected under the Trade Secrets Directive (EU) 2016/ 943, which allows the disclosure of trade secrets to expose misconduct, wrongdoing, or illegal activity,

provided the disclosure serves the public interest. Furthermore, according to the case law of the European Court of Human Rights (ECtHR), public disclosures may be protected when the public interest in the disclosed information outweighs other considerations, even if the revealed misconduct is not formally illegal and the whistleblower is subject to strict confidentiality obligations. Signatories are encouraged to refrain from harassment or retaliation when a whistleblower acts in genuine belief that the public disclosure serves the public interest. (Clarification and minor extension of scope)

Explanation: Protecting the right to publicly disclose wrongdoing in AI development when it is in the public interest is crucial for democratic oversight of this powerful technology and enables citizens to shape the development and deployment of AI, ensuring it serves the common good. This is paramount given the profound societal impact of AI systems. Transparency and accountability are essential to ensure that AI development aligns with public values and avoids unintended consequences.

Signatories are encouraged to:

- Develop joint minimum standards for whistleblowing policies that address the specific risks and challenges associated with Artificial Intelligence (AI).
- Create joint standards for evaluating the effectiveness of their whistleblowing policies. This evaluation should include independent audits conducted by qualified external organizations, appointed by the AI Office, and involving the employees concerned.
- Publish whistleblowing policies along with key findings and summary recommendations from the policies evaluations.

(**Major** extension of scope)

Explanation: Common standards promote consistent and accessible reporting channels, making it easier for whistleblowers to report concerns. The ISO 37002 Whistleblowing Management Systems guidelines serve as a valuable framework. Regular evaluations help verify the system's usability and adoption by employees. Subjecting whistleblowing policies to independent audits and publicly sharing key findings demonstrates a strong commitment to ethical practices and good governance.

Signatories are encouraged to explore the establishment of joint ombudsperson services or independent advice centers to provide guidance and support to potential whistleblowers within the AI sector. These services could also serve as a point of contact for inquiries regarding potential risks or concerns of a more general or systemic nature. Signatories commit to informing people working for them about the existence of this channel and its purpose. (Extension of scope)

Explanation: Measure 10.8 highlights the importance of sharing tools, expertise, and best practices across the AI ecosystem. Establishing joint whistleblowing services, such as ombudspersons or independent advice centers, aligns closely with this principle and the commitments outlined in the Code of Practice. These centralized services would provide essential guidance and support to potential whistleblowers while enabling individuals working for the signatories to ensure that broader or systemic risks in the AI sector are effectively addressed.)

Signatories are encouraged to permit employees to seek independent advice and assistance from third-party organizations specializing in whistleblowing or ethical considerations. Clear guidelines should be established regarding the scope and conditions of such external advice. A list of potential third-party organizations may be considered to assist employees in their decision-making. (**Major** extension of scope)

Explanation: Access to independent advice can help to reduce barriers to reporting by providing employees with a safe and confidential space to explore their concerns and receive guidance from experts before deciding whether to proceed with a formal report.

Signatories with less than 50 employees are encouraged to implement the whistleblower protections outlined above, in particular, at least annually informing employees of an AI Office mailbox, if it is operational, and implementing measures and policies to protect whistleblowers from all forms of retaliation, including dismissal, demotion, or harassment. Smaller organizations may collaborate with other employers to establish and operate joint reporting channels. Alternatively, they can mandate the ombudsperson of the industrial organizations proposed above to fulfill these responsibilities. (Clarification)
Explanation: The financial and administrative burden associated with operating an independent reporting channel can be discouraging for small and medium-sized enterprises (SMEs). Sharing resources through joint reporting channels helps to alleviate these concerns and creates an additional incentive for establishing an ombudsperson at the industry organization level.

Potential Key Performance Indicators

KPI 18.7: Evidence of the regular evaluation of the whistleblowing policy's effectiveness by an independent third-party organization, appointed by the AI Office, with key findings and summary recommendations published. Evaluations should include:

- Stakeholder Surveys and Interviews: Measuring awareness, trust in the system, and knowledge of rights and procedures.
- Channel Feedback Analysis: Reviewing reported concerns, response timeliness, and issue resolution.
- Employee Outcomes Monitoring: Tracking whistleblower departures, reasons for leaving, and post-investigation impacts.
- Corrective Action Review: Assessing the proportion of reports leading to actions and their effectiveness.

(**Major** extension of scope)
Explanation: Individuals will only use whistleblowing channels if they feel adequately protected, believe their reports will have an impact, and trust the people receiving their reports. Evaluations help assess whether the organization's whistleblowing framework is achieving these goals, identifying gaps in protection, trust, and organizational culture, and providing actionable insights for improvement.

About Whistleblower-Netzwerk

Whistleblower-Netzwerk (WBN) works to enhance the legal protection and societal recognition of whistleblowers. Founded in 2006, this German non-profit organization provides expert opinions, position papers, and recommendations for improving whistleblower laws. WBN has contributed to the drafting of German whistleblower protection laws and has participated as an expert in public hearings of the Bundestag Legal Committee. WBN raises awareness about the importance of whistleblowers for democracy through presentations, workshops, and a traveling exhibition. The organization also offers counseling services to whistleblowers and consults with businesses, authorities, and NGOs on optimizing whistleblowing systems. WBN's extensive network includes renowned jurists, former compliance officers, and civil society representatives. such as former ECJ judge Ninon Colneric, information freedom advocate Arne Semsrott, and Matthias Spielkamp, co-founder and managing director of AlgorithmWatch.

About OAISIS

The goal of the independent non-profit project OAISIS is to support concerned individuals at the frontier of AI experts in engaging with critical developments in the public interest. OAISIS promotes responsible innovation by providing access to resources, expert assessments, and networks. Leading consultants and employees from the fields of AI, law, and journalism support the work of OAISIS.

Contact:

Kosmas Zittel, Managing Director
zittel@whistleblower-net.de
Phone/Signal: +49 176 84915150
https://whistleblower-net.de/

Karl Koch, Co-Founder of OAISIS
karl.koch@oais.is
Phone/Signal: +4915227434987
https://oais.is/
https://third-opinion.org/